

Тема 6. Характеристики систем пам'яті. Ієрархія запам'ятовуючих пристроїв. Принципи побудови основних типів пам'яті

У будь-якій ОМ, незалежно від її архітектури, програми і дані зберігаються в пам'яті. Функції пам'яті забезпечуються запам'ятовуючими пристроями (ЗП), призначеними для фіксації, зберігання і видачі інформації в процесі роботи ОМ. Процес фіксації інформації в ЗП називається записом, процес видачі інформації – читанням або зчитуванням, а спільно їх визначають як процеси звернення до ЗП.

6.1 Характеристики систем пам'яті

Перелік основних характеристик, які необхідно враховувати, розглядаючи конкретний вид ЗП, включає:

- місце розташування;
- ємність;
- одиницю пересилки;
- метод доступу; швидкодію;
- фізичний тип;
- фізичні особливості;
- вартість.

За місцем розташування ЗП розділяють на процесорні, внутрішні і зовнішні. Найбільш швидкісні види пам'яті (реєстри, кеш-пам'ять першого рівня) зазвичай розміщують на загальному кристалі з центральним процесором, а реєстри загального призначення взагалі вважаються частиною ЦП. Другу групу (внутрішню пам'ять) утворюють ЗП, розташовані на системній платі. До внутрішньої пам'яті відносять основну пам'ять, а також кеш-пам'ять другого і подальших рівнів (кеш-пам'ять другого рівня може також розміщуватися на кристалі процесора). Повільні ЗП великої ємності (магнітні і оптичні диски, магнітні стрічки) називають зовнішньою пам'яттю, оскільки до ядра ОМ вони підключаються аналогічно пристроям вводу/виводу.

Ємність ЗП характеризують числом бітів або байтів, яке може зберігатися в запам'ятовуючому пристрої. На практиці застосовуються крупніші одиниці, а для їх позначення до слів «біт» або «байт» додають префікси: кіло, мега, гіга, тера, пета, екса (kilo, mega, giga, tera, peta, exa). Стандартно ці префікси означають множення основної одиниці вимірювань на 10^3 , 10^6 , 10^9 , 10^{12} , 10^{15} і 10^{18} відповідно. У обчислювальній техніці, орієнтованій на двійкову систему числення, вони відповідають значенням, достатньо близьким до стандартних, але таких, що є цілим ступенем числа 2, тобто 2^{10} , 2^{20} , 2^{30} , 2^{40} , 2^{50} , 2^{60} .

Важливою характеристикою ЗП є одиниця пересилки. Для основної пам'яті (ОП) одиниця пересилки визначається шириною шини даних, тобто кількістю бітів, які передаються по лініях шини паралельно. Зазвичай одиниця пересилки рівна довжині слова, але не обов'язково. Стосовно зовнішньої пам'яті дані часто передаються одиницями, що перевищують розмір слова, і такі одиниці називаються блоками.

Під час оцінки швидкодії необхідно враховувати вживаний в даному типі ЗП метод доступу до даних. Розрізняють чотири основні методи доступу:

- Послідовний доступ. ЗП з послідовним доступом орієнтований на зберігання інформації у вигляді послідовності блоків даних, що називаються записами. Для доступу до потрібного елемента (слова або байта) необхідно прочитати всі передуючі йому дані. Час доступу залежить від положення необхідного запису в послідовності записів на носіїві інформації і позиції елемента всередині даного запису. Прикладом може служити ЗП на магнітній стрічці.
- Прямий доступ. Кожен запис має унікальну адресу, що відображає її фізичне розміщення на носіїві інформації. Звернення здійснюється як адресний доступ на початок запису, з подальшим послідовним доступом до певної одиниці інформації всередині запису. У результаті час доступу до певної позиції є величиною змінною. Такий режим характерний для магнітних дисків.

- Довільний доступ. Кожен елемент пам'яті має унікальну фізичну адресу. Звернення до будь-якої комірки займає один і той же час і може проводитись у довільній черговості. Прикладом можуть служити запам'ятовуючі пристрої основної пам'яті.
- Асоціативний доступ. Цей вид доступу дозволяє виконувати пошук комірок, що містять таку інформацію, в якій значення окремих бітів збігається зі станом однойменних бітів у заданому зразку. Порівняння здійснюється паралельно для всіх елементів пам'яті, незалежно від її ємності. За асоціативним принципом побудовані деякі блоки кеш-пам'яті.

Швидкодія ЗП є одним з найважливіших його показників. Для кількісної оцінки швидкодії зазвичай використовують три параметри:

- Час доступу (ТД). Для пам'яті з довільним доступом він відповідає інтервалу часу від моменту надходження адреси до моменту, коли дані заносяться в пам'ять або стають доступними. У ЗП з рухомим носієм інформації – це час, що витрачається на встановлення головки запису/зчитування (або носія) в потрібну позицію.
- Тривалість циклу пам'яті або період звернення (Тц). Поняття застосовується до пам'яті з довільним доступом, для якої воно означає мінімальний час між двома послідовними зверненнями до пам'яті. Період звернення включає час доступу плюс деякий додатковий час. Додатковий час може бути потрібним для затухання сигналів на лініях, а в деяких типах ЗП, де зчитування інформації приводить до її руйнування, – для відновлення зчитаної інформації.
- Швидкість передачі. Це швидкість, з якою дані можуть передаватися в пам'ять або з неї. Для пам'яті з довільним доступом вона дорівнює $1/T_c$. Для інших видів пам'яті швидкість передачі визначається співвідношенням

$$T_N = T_A + \frac{N}{R}$$

де T_N – середній час зчитування або запису N бітів;

T_A – середній час доступу;

R – швидкість пересилки в бітах за секунду.

Кажучи про фізичний тип запам'ятовуючого пристрою, необхідно згадати три найбільш поширених технології ЗП – це напівпровідникова пам'ять, пам'ять з магнітним носієм інформації, використовувана в магнітних дисках і стрічках, і пам'ять з оптичним носієм – оптичні диски.

Залежно від застосованої технології слід враховувати і ряд фізичних особливостей ЗП, наприклад енергозалежність. В енергозалежній пам'яті інформація може бути спотворена або втрачена під час відключення джерела живлення. В енергонезалежних ЗП записана інформація зберігається і у разі відключення напруги живлення. Магнітна і оптична пам'ять – енергонезалежні.

Напівпровідникова пам'ять може бути як енергозалежною, так і ні, залежно від її типу. Крім енергозалежності потрібно враховувати, приводить зчитана інформація до її руйнування чи ні.

Вартість ЗП прийнято оцінювати відношенням загальної вартості ЗП до його ємності в бітах, тобто вартістю зберігання одного біта інформації.

6.2 Ієрархія запам'ятовуючих пристроїв

Під час створення системи пам'яті постійно доводиться вирішувати задачу забезпечення необхідної ємності і високої швидкодії за прийнятну ціну.

Найбільш поширеним підходом тут є побудова системи пам'яті ОМ за ієрархічним принципом. Ієрархічна пам'ять складається із ЗП різних типів (рис. 6.1), які, залежно від характеристик, відносять до певного рівня ієрархії.

Вищий рівень менше по ємності, швидше і має велику вартість в перерахунку на біт, ніж нижчий рівень.

Рівні ієрархії взаємозв'язані: всі дані на одному рівні можуть бути також знайдені на нижчому рівні, і всі дані на цьому нижчому рівні можуть бути знайдені на наступному рівні, що знаходиться нижче, і так далі.

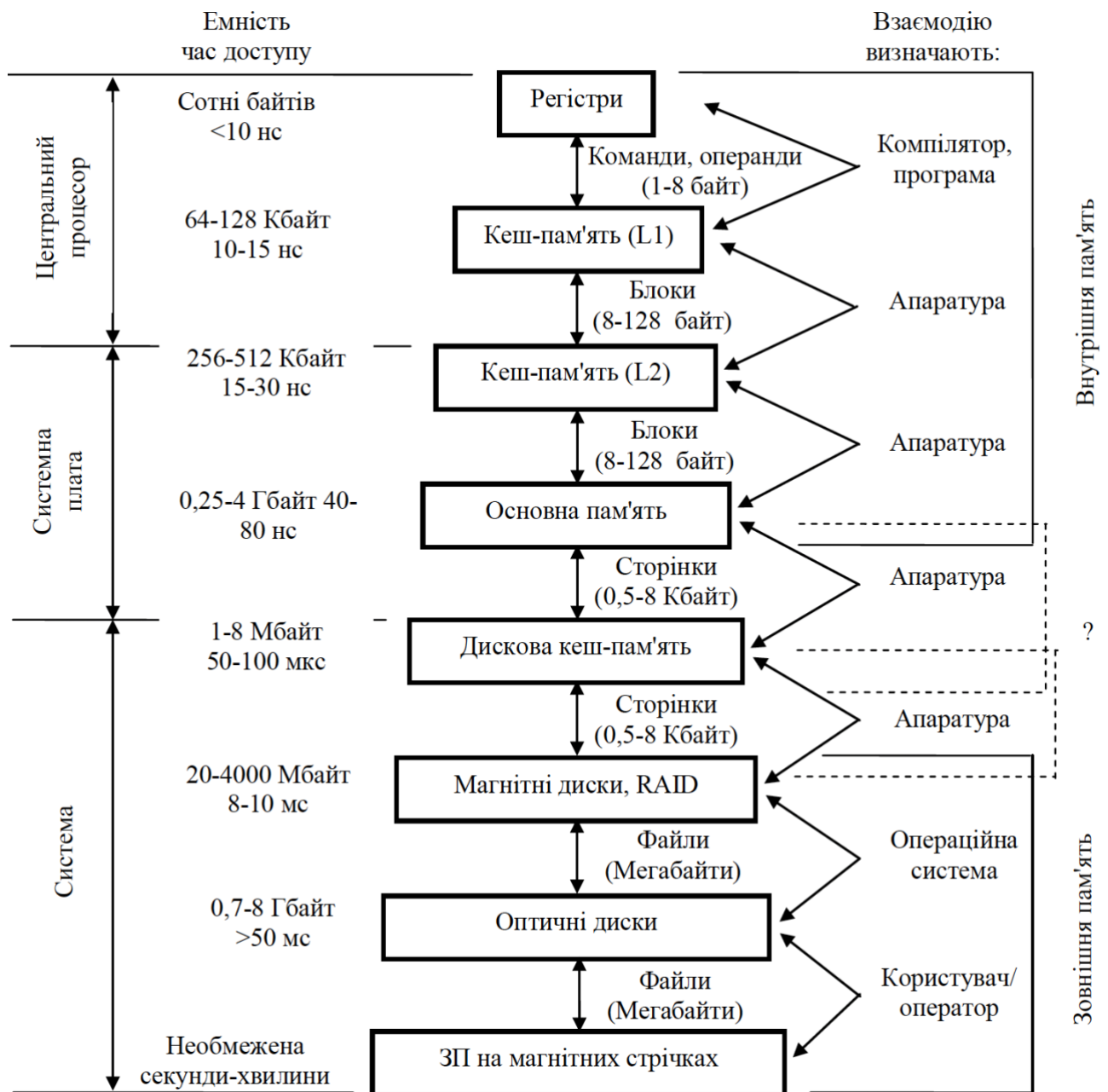


Рисунок 6.1 – Ієрархія запам'ятовуючих пристроїв

На кожному рівні ієрархії інформація розбивається на блоки, які є найменшою інформаційною одиницею, що пересилається між двома сусідніми рівнями ієрархії. Розмір блоків може бути фіксованим або змінним. У разі фіксованого розміру блока ємність пам'яті зазвичай кратна його розміру. Розмір блоків на кожному рівні ієрархії найчастіше різний і збільшується від верхніх рівнів до нижніх.

Під час доступу до команд і даних, наприклад для їх зчитування, спочатку проводиться пошук у пам'яті верхнього рівня. Факт виявлення потрібної

інформації називають попаданням (hit), інакше говорять про промах (miss). У разі промаху проводиться пошук в ЗП наступного нижчого рівня, де також можливі попадання або промах. Після виявлення необхідної інформації виконується послідовна пересилка блока, що містить шукану інформацію, з нижніх рівнів на верхні. Слід зазначити, що незалежно від числа рівнів ієрархії пересилка інформації може здійснюватися тільки між двома сусідніми рівнями.

Найшвидший, але і мінімальний по ємності тип пам'яті – це внутрішні регістри ЦП, які іноді об'єднують поняттям надоперативний запам'ятовуючий пристрій, – НОЗП. Як правило, кількість регістрів невелика, хоча в архітектурі зі скороченим набором команд їх число може доходити до декількох сотень. Основна пам'ять (ОП), значно більшої ємності, розташовується декількома рівнями нижче. Між регістрами ЦП і основною пам'яттю часто розміщують кеш-пам'ять, яка за ємністю відчутно програє ОП, але істотно перевершує останню за швидкістю, поступаючись в той же час НОЗП. У більшості сучасних ОМ є декілька рівнів кеш-пам'яті, які позначають буквою L і номером рівня кеш-пам'яті. На рис. 3.1 показані два таких рівні. В останніх розробках все частіше з'являється також третій рівень кеш-пам'яті (L3), причому розробники ОМ говорять про доцільність введення і четвертого рівня – L4. Кожен подальший рівень кеш-пам'яті має більшу ємність, але одночасно і меншу швидкість в порівнянні з попередньою. Всі види внутрішньої пам'яті реалізуються на основі напівпровідникових технологій і в основному є енергозалежними. Довготривале зберігання великих об'ємів інформації (програм і даних) забезпечується зовнішніми ЗП, серед яких найбільш поширені запам'ятовуючі пристрої на базі магнітних і оптичних дисків, а також ЗП на магнітних стрічках.

Нарешті, ще один рівень ієрархії може бути доданий між основною пам'яттю і дисками. Цей рівень носить назву дискової кеш-пам'яті і реалізується у вигляді самостійного ЗП, такого, що включається до складу магнітного диска. Дискова кеш-пам'ять істотно покращує продуктивність при обміні інформацією між дисками і основною пам'яттю.

Ієрархія може бути доповнена і іншими видами пам'яті. Так, деякі моделі ОМ фірми ІВМ включають так звану розширену пам'ять (expanded storage), виконану на основі напівпровідникової технології, але маючу меншу швидкодію і вартість у порівнянні з ОП. Строго кажучи, цей вид пам'яті не входить в ієрархію, а є відгалуженням від неї, оскільки дані можуть передаватися тільки між розширеною і основною пам'яттю, але не допускається обмін між розширеною і зовнішньою пам'яттю.

6.3 Принципи побудови основних типів пам'яті

Способи організації пам'яті залежать від методів розташування та пошуку інформації в запам'ятовуючому масиві. За цією ознакою розрізняють адресну та безадресну пам'яті.

6.3.1 Адресні запам'ятовуючі пристрої

В пам'яті з адресною організацією розташування та пошук інформації в ЗП оснований на використанні адреси зберігання слова (числа, команди і т.п.). Адресою служить номер комірки ЗП, в якій це слово розташовується.

Під час запису (або зчитування) слова в ЗП команда, яка ініціює цю операцію, повинна вказувати адресу (номер комірки), за якою проводиться запис (зчитування).

Типова структура адресної пам'яті показана на рис. 6.2.

Структура містить запам'ятовуючий масив (ЗМ) з N n -розрядних комірок та його апаратне обладнання, яке включає в себе регістр адреси (РГА), який має k розрядів ($k \geq \log_2 N$), інформаційний регістр (РГІ), блок адресної вибірки (БАВ), блок підсилювачів зчитування (БПЗ), блок розрядних підсилювачів-формуваців сигналів запису (БПФЗ) та блок управління пам'яттю (БУП).

По коду адреси в РГА БАВ формує у відповідній комірці пам'яті сигнали, які дозволяють провести зчитування або запис слова. Цикл звернення до пам'яті ініціюється надходженням в БУП ззовні сигналу звернення.

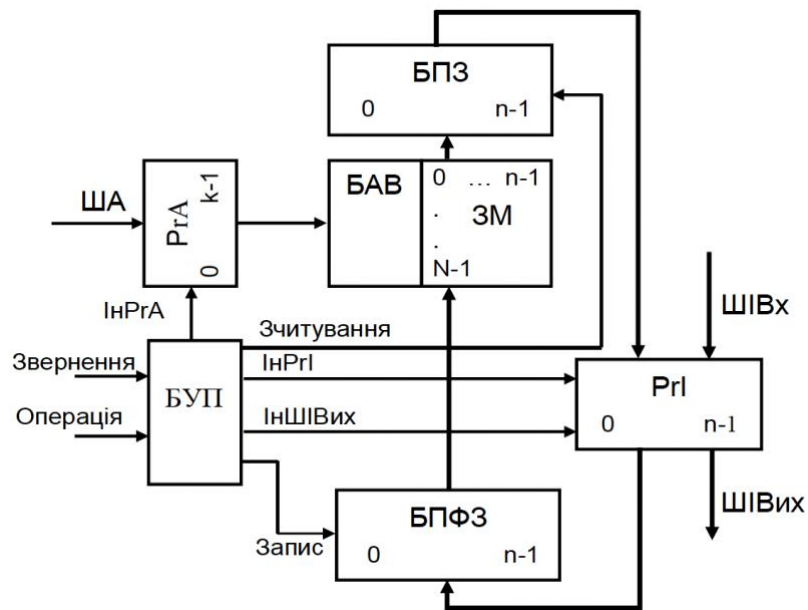


Рисунок 6.2 – Структура адресної пам'яті

Загальна частина циклу звертання включає в себе прийом в PrgA з шини ША адреси звертання та прийом в БУП і розшифровку управляючого сигналу операція, який вказує вид запитуваної операції (зчитування або запис).

Далі під час зчитування БАП дешифрує адресу, виконує вибірку комірки пам'яті, посилає сигнал зчитування у вибрану комірку ЗП. Код записаного в комірку слова зчитується, підсилюється підсилювачами зчитування та передається в ПрІ. Операція зчитування завершується видачею слова з ПрІ на вихідну інформаційну шину ШІВих.

Під час запису, окрім виконання вказаної вище загальної частини циклу звертання, проводиться прийом слова, яке записується з вхідної інформаційної шини ШІВх в ПрІ. Потім у комірку, яку вибрав БАП, записується слово з ПрІ. Зазвичай роль вхідного і вихідного виконує один і той же регістр.

Підсилювачі зчитування/запису служать для електричного узгодження сигналів на лініях даних і внутрішніх сигналів ІМС. Зазвичай число підсилювачів дорівнює числу запам'ятовуючих елементів у рядку матриці, і всі вони під час звернення до пам'яті підключаються до вибраної горизонтальної лінії. Кожна група підсилювачів, яка створює комірку, підключена до відповідних стовпців

матриці, тобто вибір потрібної комірки в рядку забезпечується активізацією вертикальних ліній.

Блок управління БУП генерує необхідні послідовності управляючих сигналів, які ініціюють роботу окремих вузлів пам'яті.

6.3.2 Безадресні запам'ятовуючі пристрої

У безадресній пам'яті для пошуку інформації в запам'ятовуючому масиві використовують не адреси комірок, а інші принципи. До цього типу пам'яті відноситься асоціативна та стекова пам'яті.

Асоціативна пам'ять. У пам'яті цього типу пошук потрібної інформації проводиться не за адресою, а за її змістом (за асоціативною ознакою). При цьому пошук за асоціативною ознакою (або послідовно за окремими розрядами цієї ознаки) відбувається паралельно у часі для всіх комірок запам'ятовуючого масиву. В багатьох випадках асоціативний пошук дозволяє істотно спростувати та прискорити обробку даних. Це досягається за рахунок того, що в пам'яті цього типу операція зчитування інформації суміщена з виконанням ряду логічних операцій.

Типова структура асоціативної пам'яті показана на рис. 6.3.

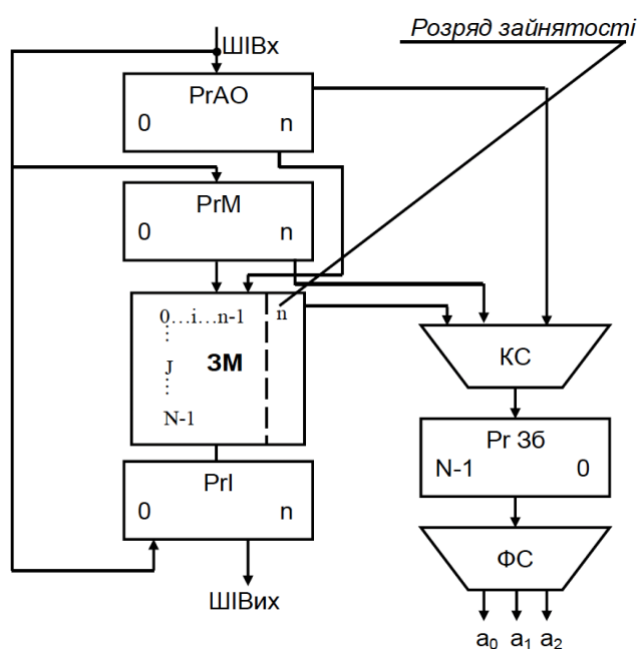


Рисунок 6.3 – Структура асоціативної пам'яті

Запам'ятовуючий масив має $N(n+1)$ -розрядних комірок. Для вказування зайнятості комірки використовують службовий n -й розряд (0 – комірка вільна, 1 – в комірці записане слово).

По вхідній інформаційній шині ШВх в реєстр асоціативної ознаки (РгАО) в розряди $0 \div n-1$ надходить n -розрядний асоціативний запит, а в реєстр маски (РгМ) – код маски пошуку, при цьому n -й розряд РгМ установлюється в 0.

Асоціативний пошук проводиться лише для сукупності розрядів РгАО, яким відповідає 1 в РгМ (не замасковані розряди РгАО). Для слів, у яких цифри в розрядах збіглися з незамаскованими розрядами РгАО, комбінаційна схема (КС) установлює 1 у відповідні розряди реєстра збігання (РгЗб) і 0 в решту розрядів.

Комбінаційна схема формування результатів асоціативного звертання (ФС) формує із слова, яке утворилося в РгЗб, сигнали a_0, a_1, a_2 , що відповідають випадкам відсутності слів у ЗМ, які задовольняють асоціативні ознаки та наявність одного (a_1) і більше (a_2) таких слів.

Формування вмісту РгЗб та сигналів a_0, a_1, a_2 , за вмістом РгАО, РгМ та ЗМ, називається операцією контролю асоціації. Ця операція є складовою частиною операцій зчитування та запису, хоч вона має самостійне значення.

Під час зчитування спочатку проводиться контроль асоціації за асоціативною ознакою в РгАО. Потім при $a_0=1$ зчитування скасовується через відсутність потрібної інформації, при $a_1=1$ зчитується слово, на яке вказує одиниця в РгЗб, при $a_2=1$ в РгІ зчитується слово із комірки, яка має найменший номер серед комірок, позначених 1 в РгЗб. Із РгІ зчитане слово видається на вихідну інформаційну шину (ШВих).

Під час запису інформації спочатку відшукується вільна комірка. Для цього виконується операція контролю асоціації при РгАО = 111...10 та РгМ = 000...01, при цьому вільні комірки позначаються 1 в РгЗб. Для запису вибирається вільна комірка з найменшим номером або та, яка довше не використовувалася. В неї записують слово, яке надійшло із ШВх в РгІ.

За допомогою операції контролю асоціації можна, не зчитуючи слів з пам'яті, визначити за вмістом РгЗб, скільки в пам'яті слів, які задовольняють асоціативні ознаки, наприклад реалізувати запити, типу скільки студентів мають відмінні оцінки з даної дисципліни. Під час використання відповідних комбінаційних схем в асоціативній пам'яті можуть використовуватись досить складні логічні операції, такі, як пошук більшого (меншого) числа та інше.

Відзначимо, що для асоціативної пам'яті необхідні запам'ятовуючі елементи, які дозволяють виконувати зчитування без знищення записаної в них інформації.

Через відносно високу вартість асоціативна пам'ять рідко використовується як самостійний вид пам'яті.

Стекова пам'ять. Стекову пам'ять можна розглядати як сукупність комірок, що утворюють одновимірний масив, у якому сусідні комірки зв'язані одна з одною розрядними ланцюгами передачі слів. Запис нового слова проводиться в верхню комірку (вершину стека – комірку 0), при цьому всі записані раніше слова (включаючи і слова, записані в комірці 0) зсуваються вниз, в сусідні комірки з більшими на 1 номерами.

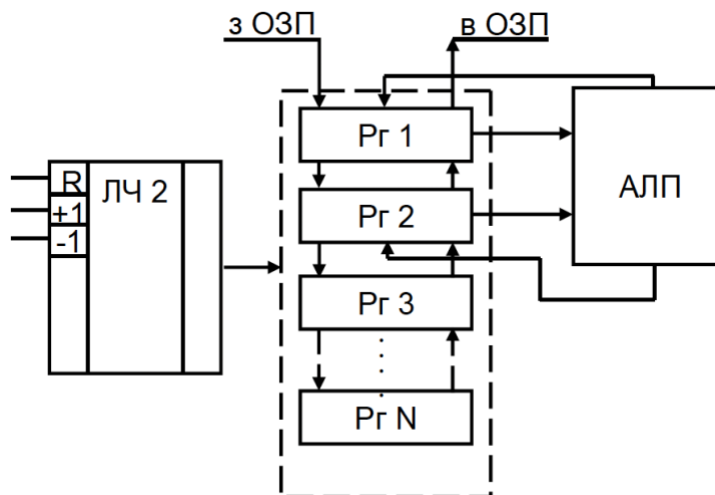
Зчитування можливе тільки з верхньої (нульової) комірки пам'яті, при цьому, якщо проводиться зчитування з видаленням, всі інші слова в пам'яті зсуваються вгору, в сусідні комірки з більшими номерами. В цій пам'яті порядок зчитування слів відповідає правилу: останнім надійшов – першим обслужився.

У ряді пристроїв розглянутого типу передбачається також операція простого зчитування слова з нульової комірки (без його видалення і зсуву слова в пам'яті). Іноді стекова пам'ять має лічильник стека (ЛчСт), який показує кількість занесених у пам'ять слів. Сигнал $ЛчСт = 0$ відповідає порожньому стеку.

Структурна схема стекової пам'яті (СтП) приведена на рис. 6.4.

Пошук і переміщення інформації в СтП здійснюється з допомогою реверсивного лічильника (Лч). Він у будь-який момент часу вказує кількість зайнятих комірок СтП. Під час запису інформації показання Лч збільшуються на

одиницю, а вміст СтП переміщується на одну комірку вниз, а при зчитуванні – показання Лч зменшується на одиницю і вміст СтП переміщується на одну комірку вгору.



Риснок 6.4 – Структурна схема стекової пам'яті

Оперативний запам'ятовуючий пристрій (ОЗП) має зв'язок тільки з верхньою коміркою пам'яті (Рг1), а процесор – з двома верхніми (Рг1 і Рг2). Операція, яка задається програмою, виконується над вмістом двох верхніх комірок – Рг1 і Рг2, а результат операції завжди залишається в Рг2. Якщо Рг1 вільний, то перед виконанням наступної операції вміст Рг2 передається в Рг1, із Рг3 в Рг2 і т.д. (вміст СтП переміщується на одну комірку вгору), а показання Лч зменшуються на одиницю.

Стекова пам'ять (її ще називають магазинна пам'ять) широко застосовується як в обчислювальних системах високої продуктивності, так і в міні- і мікро-ЕОМ, і пропонує використання безадресних команд обробки масивів даних, які утворюють стек.